



globus online

Enabling Your Campus to Simplify Research Data Management with Globus Online

Steve Tuecke

Raj Kettimuthu

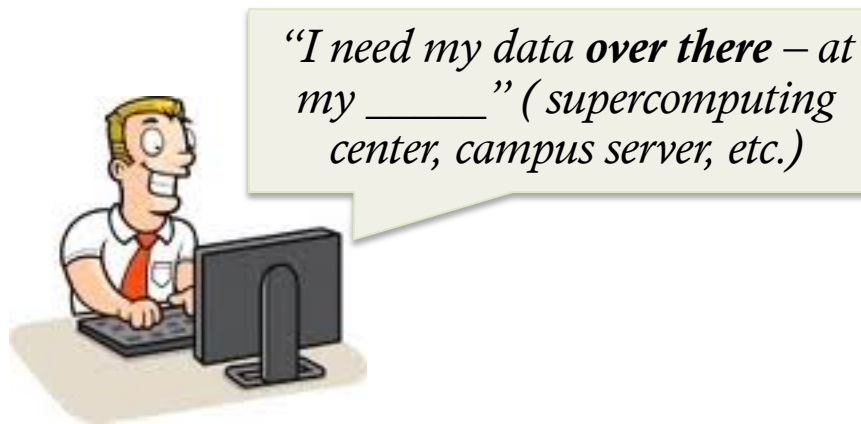
XSEDE 2012

July 16, 2012



The Challenge: Moving Big Data *Easily*

- What should be trivial...



- ... can be painfully tedious and time-consuming





What is Globus Online?

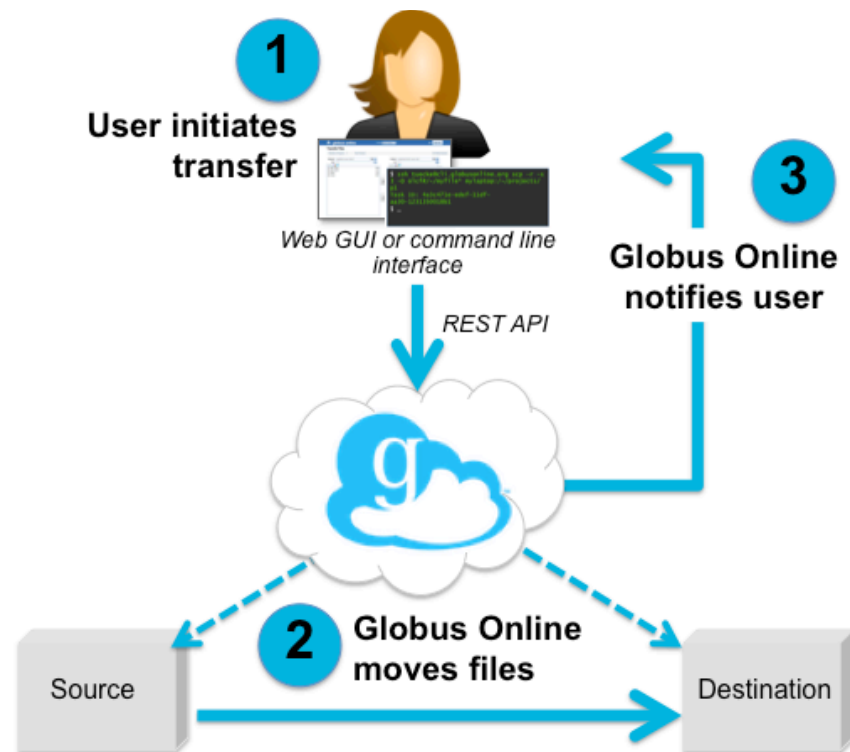
- **Reliable file transfer.**

- Easy “fire-and-forget” transfers
- Automatic fault recovery
- High performance
- Across multiple security domains

- **No IT required.**

- Software as a Service (SaaS)
 - No client software installation
 - New features automatically available
- Consolidated support & troubleshooting
- Works with existing GridFTP servers
- Globus Connect solves “last mile problem”

- **Recommended by XSEDE, Blue Waters, NERSC, ALCF, Advanced Photon Source, many Universities**





globus online



Globus Online Demo



Logging into the CLI

Interactive login to command line interface:

```
$ ssh tuecke@cli.globusonline.org
```

Running commands remotely:

```
$ ssh tuecke@cli.globusonline.org <command>
```

```
$ ssh tuecke@cli.globusonline.org scp -r -s 3 -D \  
    nersc#dtn:~/myfile* mylaptop:~/projects/p1  
Task ID: 4a3c471e-edef-11df-aa30-1231350018b1  
$ _
```

Using CLI with gsissh:

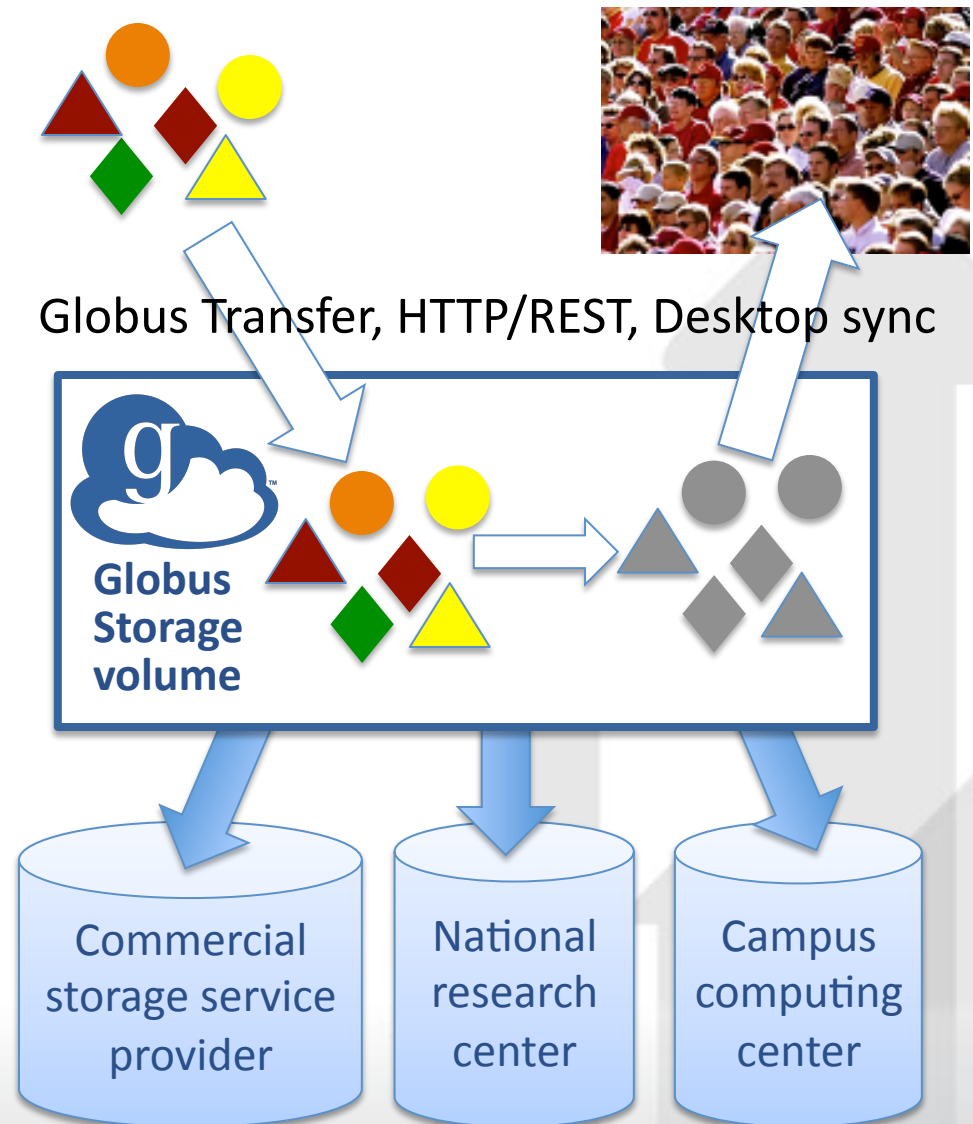
```
$ gsissh tuecke@cli.globusonline.org <command>
```



Alpha

Globus Storage: For when you want to ...

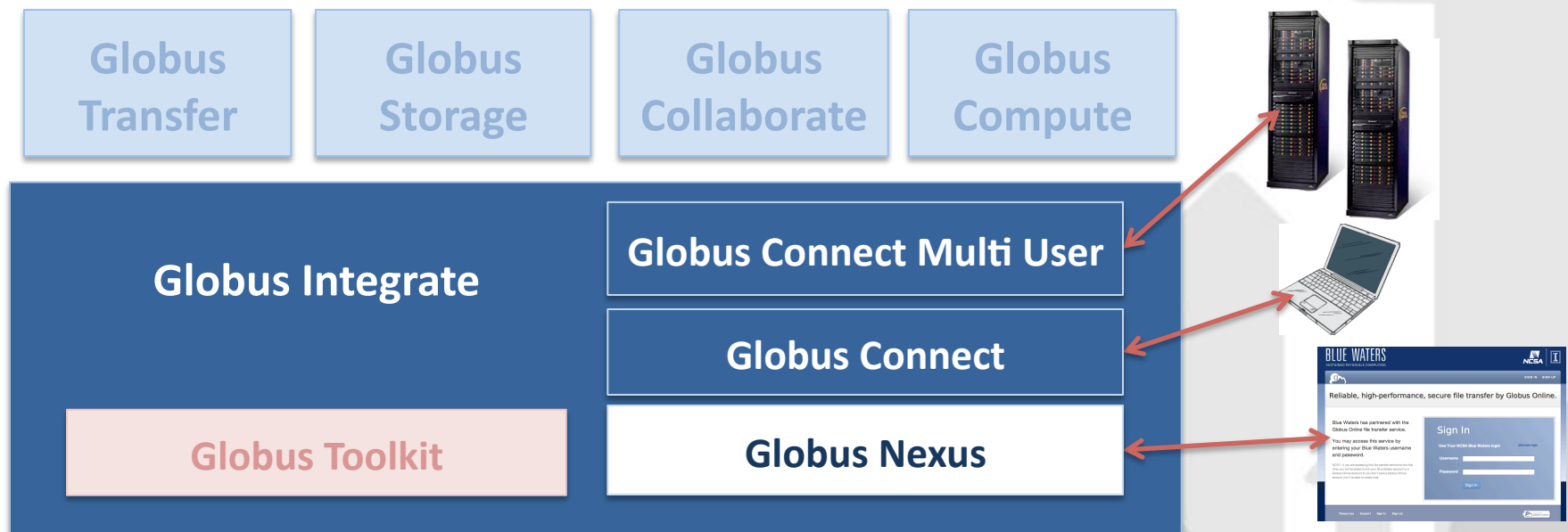
- **Place** your data where you want
- **Access** it from anywhere via different protocols
- **Update** it, **version** it, and take **snapshots**
- **Share** versions with who you want
- **Synchronize** among locations





Globus Integrate: For when you want to...

- Integrate with the Globus research cloud ecosystem
- Write programs that access/manage:
 - user identities, profiles, groups, resources
 - data, compute and collaboration



... via REST APIs and command line programs



For More Information

- **Visit <https://www.globusonline.org/signup> to:**
 - Get a free account and start moving files
- **Visit www.globusonline.org for:**
 - Tutorials, FAQs, Pro Tips, Troubleshooting
 - Papers
 - Case Studies
- **Contact support@globusonline.org for:**
 - Help getting started
 - Help using the service
- **Follow us at [@globusonline](https://twitter.com/globusonline) on Twitter and [Globus Online](https://www.facebook.com/globusonline) on Facebook**



Hands-on Session

The goal of this session is to show you (hands-on) how to take a resource and turn it into a GO endpoint.

Each of you is provided with an amazon EC2 machine for this tutorial.



Log into your host

Your slip of paper has the host information.

Log in as user “xsede12”:

ssh xsede12@ec2-xx-xx-xx-xx.compute-1.amazonaws.com

Use the password on the slip of paper.

xsede12 has passwordless sudo privileges.



Setup SSH keys

Upload xsede12's SSH key to Globus Online (/home/xsede12/.ssh/id_rsa.pub).

Each host also has five users: joe, bob, sue, ann, and sam.

Password for joe, bob, sue, ann: the same password as the “xsede12” user.

Password for sam is “xsede12tutorial” for all hosts.



Globus Connect Multi-User

- **What is GCMU?**

- Multi-user version of Globus Connect
- Packages a GridFTP server and MyProxy CA, pre-configured for use with Globus Online

- **Why GCMU?**

- Create transfer endpoints in minutes
- Avoid complex GridFTP install

- **To download:** <https://www.globusonline.org/gcmu/>



"We used GCMU to form a campus-wide GSI authentication service spanning multiple servers. Now my users have a fast, easy way to get their data wherever it needs to go, and the setup process was trivial."

--University of Michigan



"As a resource admin, I've found GCMU an exceedingly useful tool.... With GCMU, setting up a GridFTP server and handling authentication for multiple users is easy."

--Oak Ridge National Lab

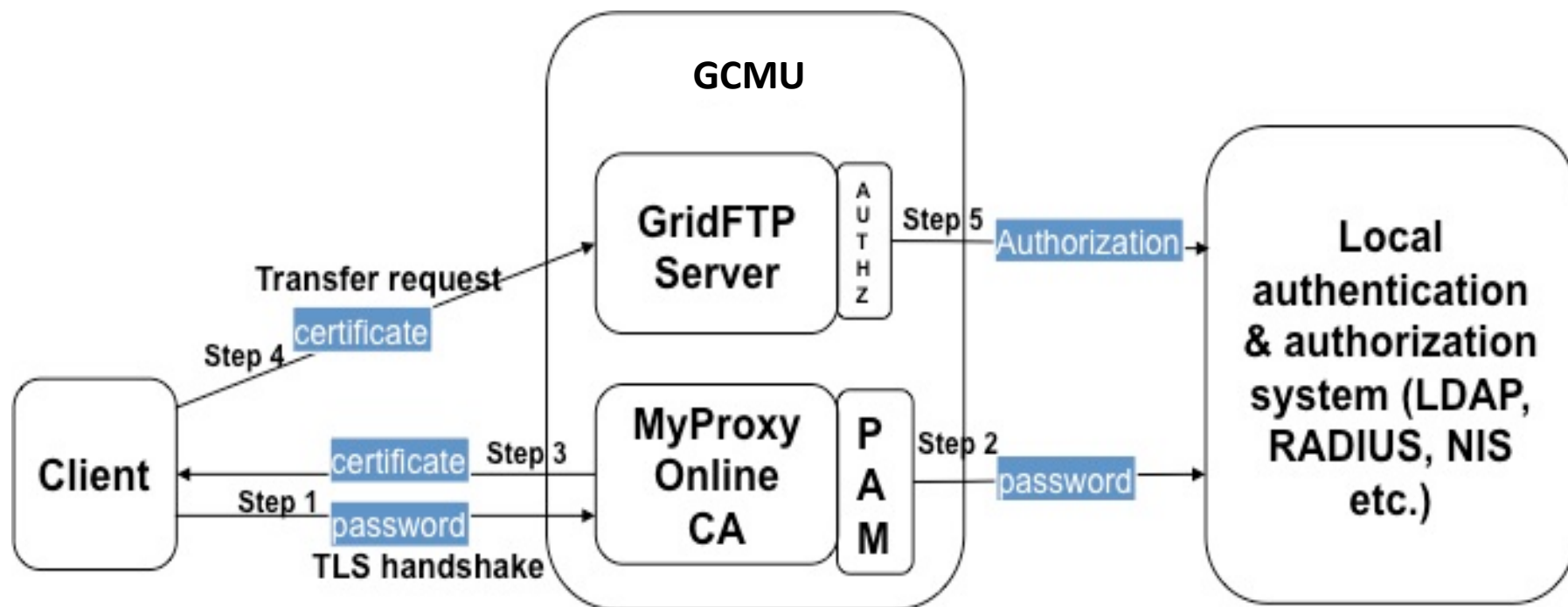


GCMU Deployments





Globus Connect Multi-User





Globus Connect Multi User

```
cd /opt
```

```
sudo wget http://connect.globusonline.org/linux/stable/  
globusconnect-multiuser-latest.tgz
```

```
sudo tar xzf globusconnect-multiuser-latest.tgz
```

```
cd gcmu*
```

```
sudo ./install
```




Try doing the following

Create a file called tutorial.txt in /home/joe

Go to the GO Web UI -> Start Transfer

Select endpoint *username#xsede12*

Activate the endpoint as user “joe” (not xsede12). You should see joe's home directory.(Remember: joe's password is the same as xsede12's)

Transfer to/from the endpoint of the person sitting next to you (activate their endpoint as user “sam”).

Does tutorial.txt show up in /home/sam in the host of the person sitting next to you?



Setup SSH keys

Upload xsede12's SSH key to Globus Online (/home/xsede12/.ssh/id_rsa.pub).

Each host also has five users: joe, bob, sue, ann, and sam.

Password for joe, bob, sue, ann: the same password as the “xsede12” user.

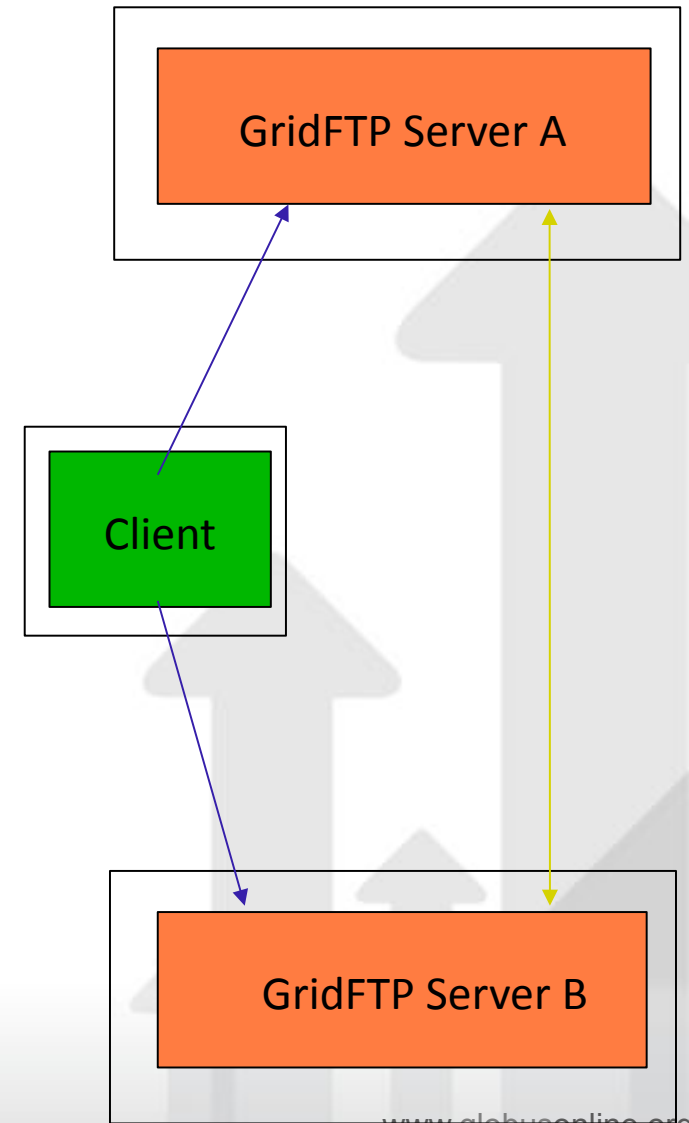
Password for sam is “xsede12tutorial” for all hosts.

- **Two channel protocol like FTP**
- **Control Channel**
 - Command/Response
 - Used to establish data channels
 - Basic file system operations eg. mkdir, delete etc
- **Data channel**
 - Pathway over which *file* is transferred
 - Many different underlying protocols can be used
 - MODE command determines the protocol



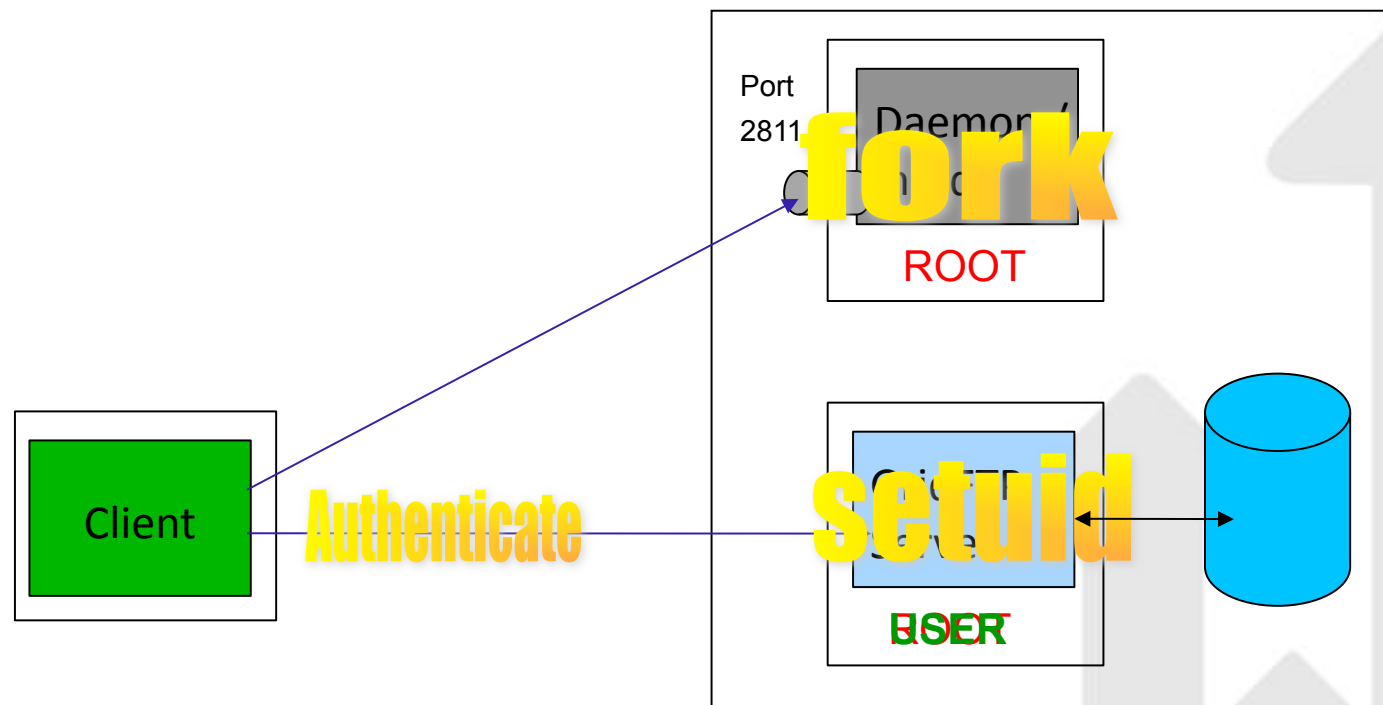
Third Party Transfer

- **Client initiates data transfer between 2 servers**
- **Client forms control channel with 2 servers.**
- **Information is routed through the client to establish data channel between the two servers.**
- **Data flows directly between servers**
 - Client is notified by each server when the transfer is complete





GridFTP Authentication and Authorization





Security Configuration

- **Installation**
- Download, untar, configure, make
- **Security configuration (server admins)**
- Obtain and install X.509 host certificate from well-known CA
- Configure trust roots
- **Security configuration (users)**
- Obtain and install user certificate from well-known CA
- Configure trust roots
- **Setup authorization (both users and admins)**
- DN to local username mapping in gridmap file
- '/DC=org/DC=doegrids/OU=People/CN=Rajkumar Kettimuthu 227852' rajk
- **Too complex for many users and small labs**

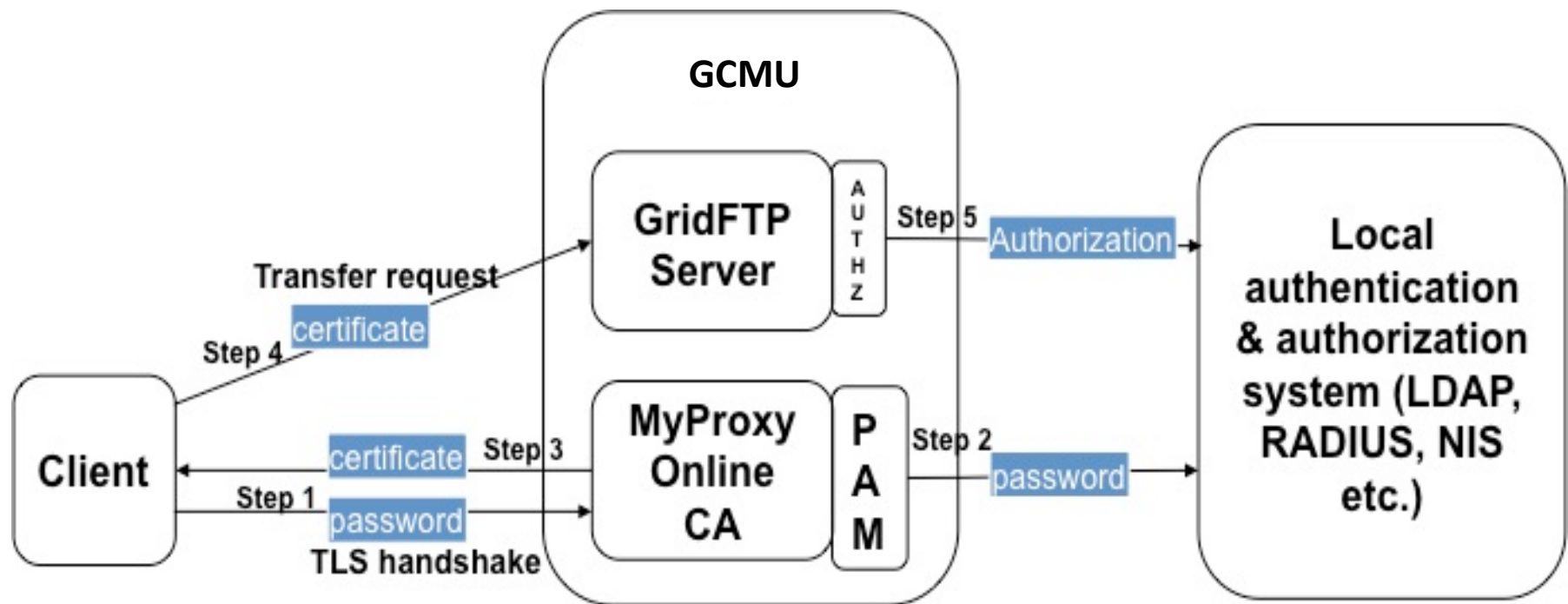


GCMU makes it trivial

- **Make GridFTP deployment trivial**
- GridFTP transfers can be achieved “instantly” even by non-experts
- **Automate the process of configuring security**
- Avoid the need for any end-user or system administrator involvement in security configuration
- **Reduce burden on both users and administrators**
- Eliminate frequent sources of errors in GridFTP configuration and use.



Globus Connect Multi-User



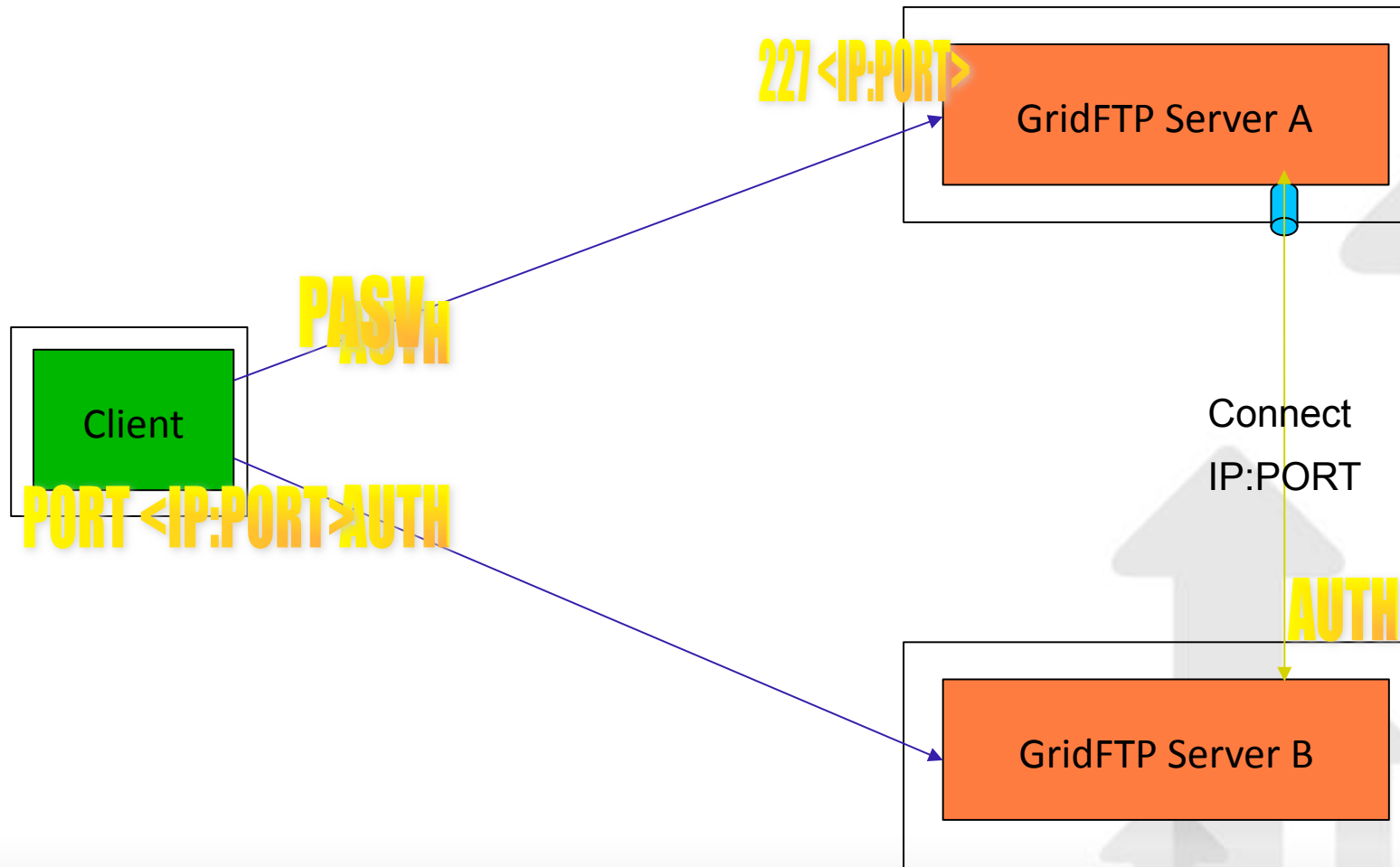


User mapping

- Username is embedded in the certificate, as this cert is used to authenticate with this site only
- “/C=US/O=Globus Consortium/OU=Globus Connect Service/CN=84591482-cba0-11e1-b791-1231381b68a7/CN=rajk”
- Authorization callout extracts username from DN, if the certificate is issued by local CA

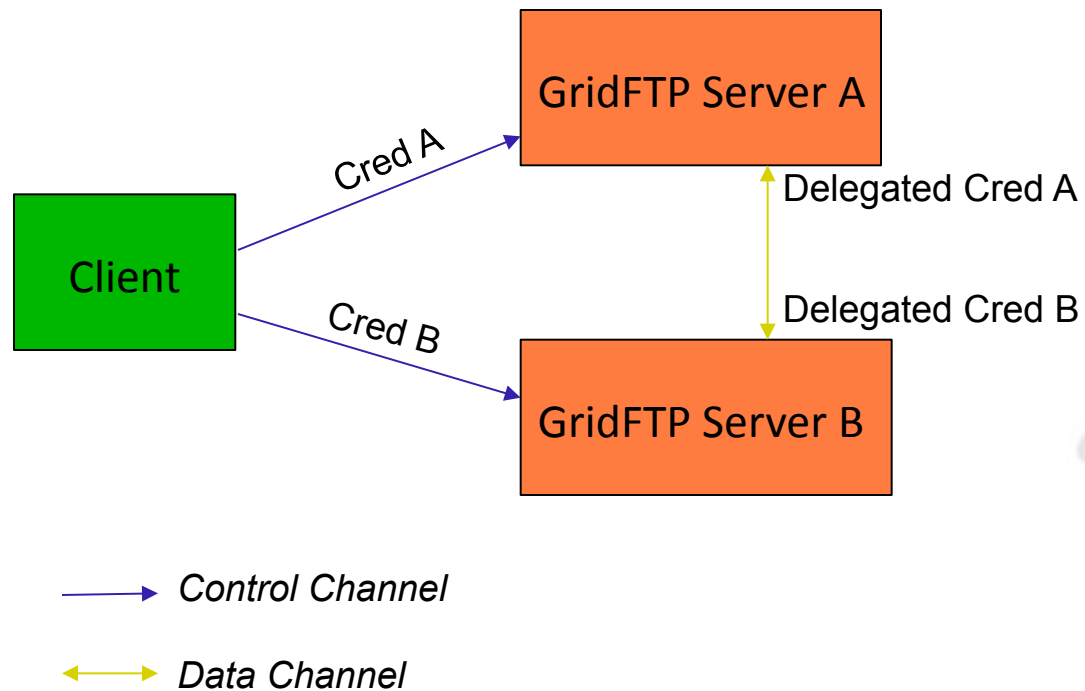


Data Channel Establishment



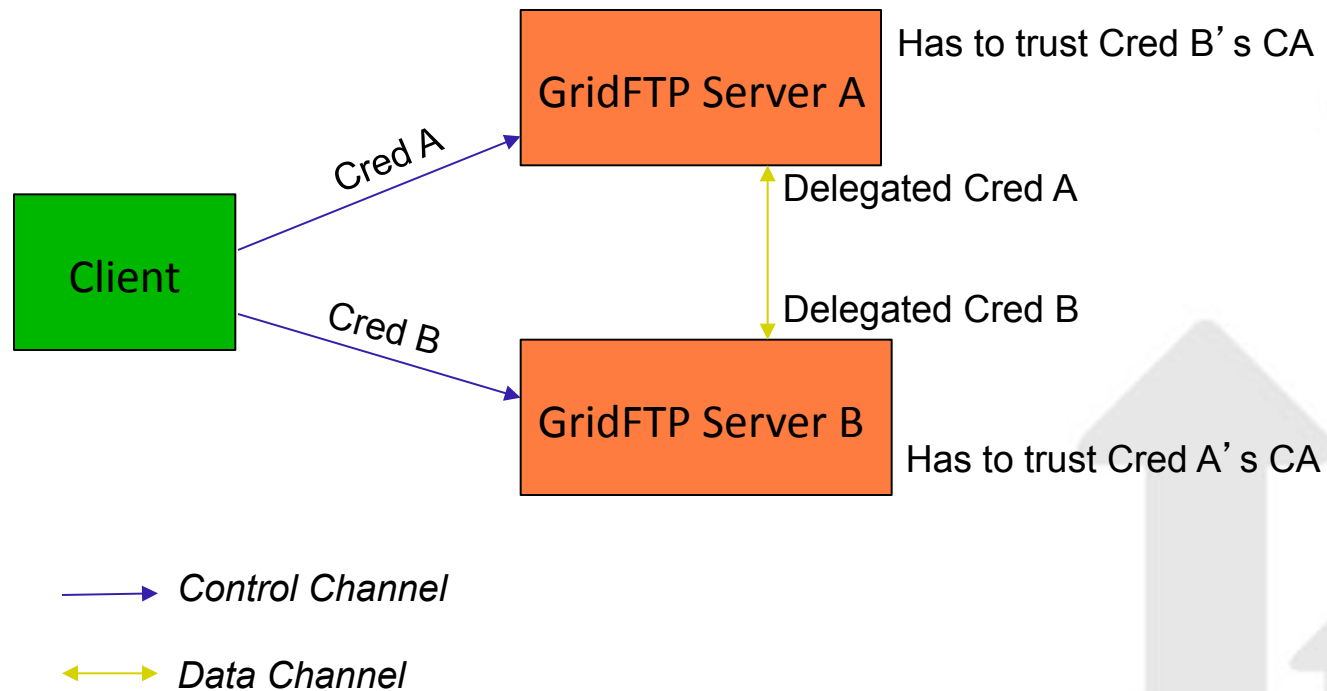


Data Channel Authentication



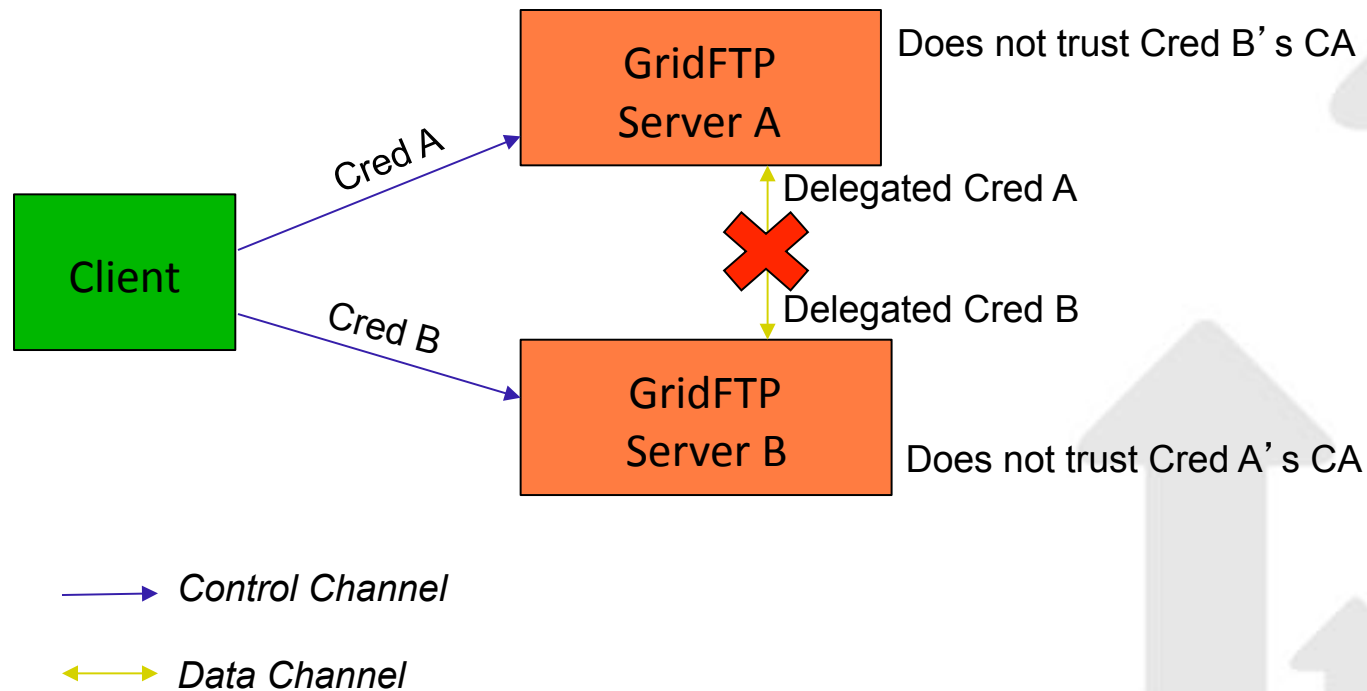


Data Channel Authentication



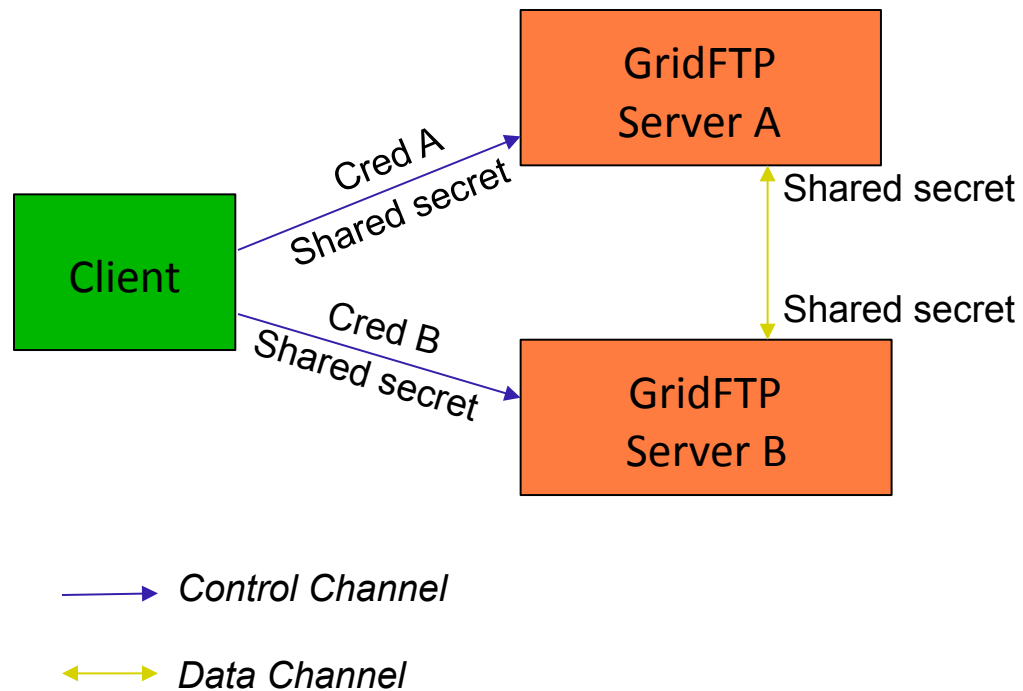


Data Channel Authentication



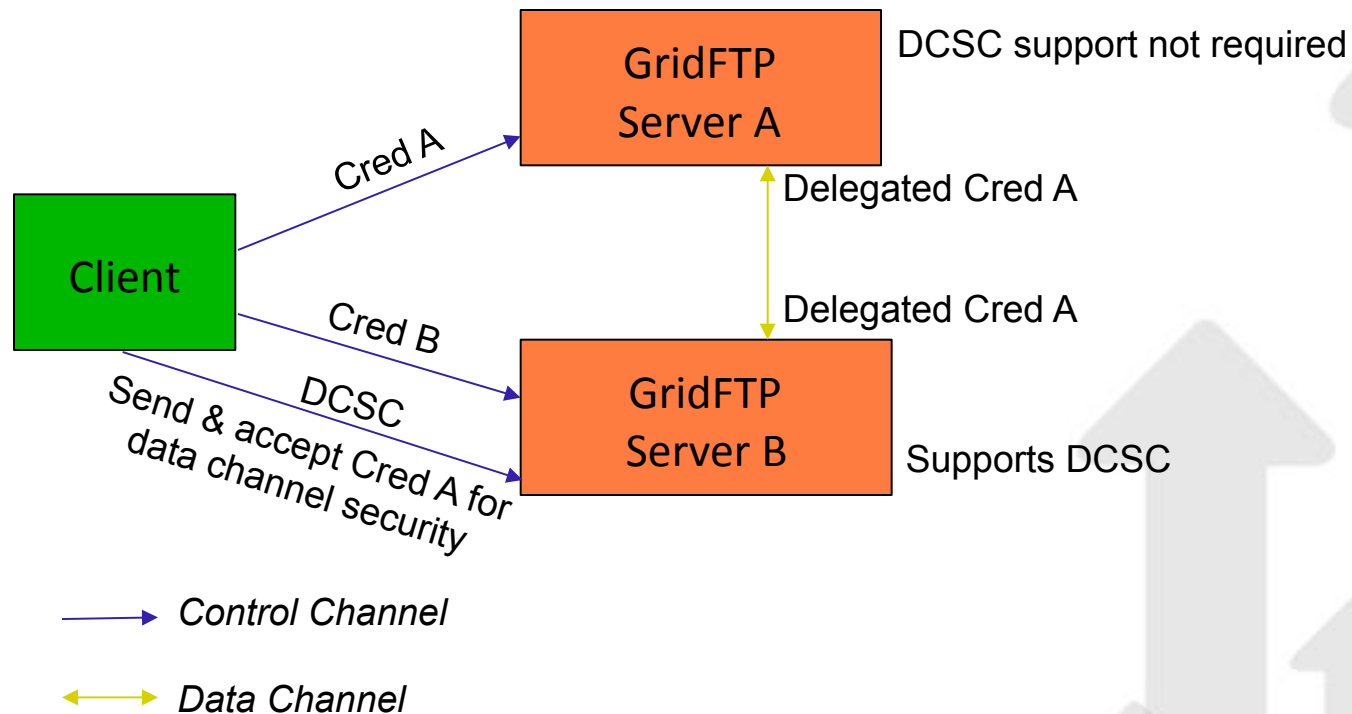


Data Channel Authentication



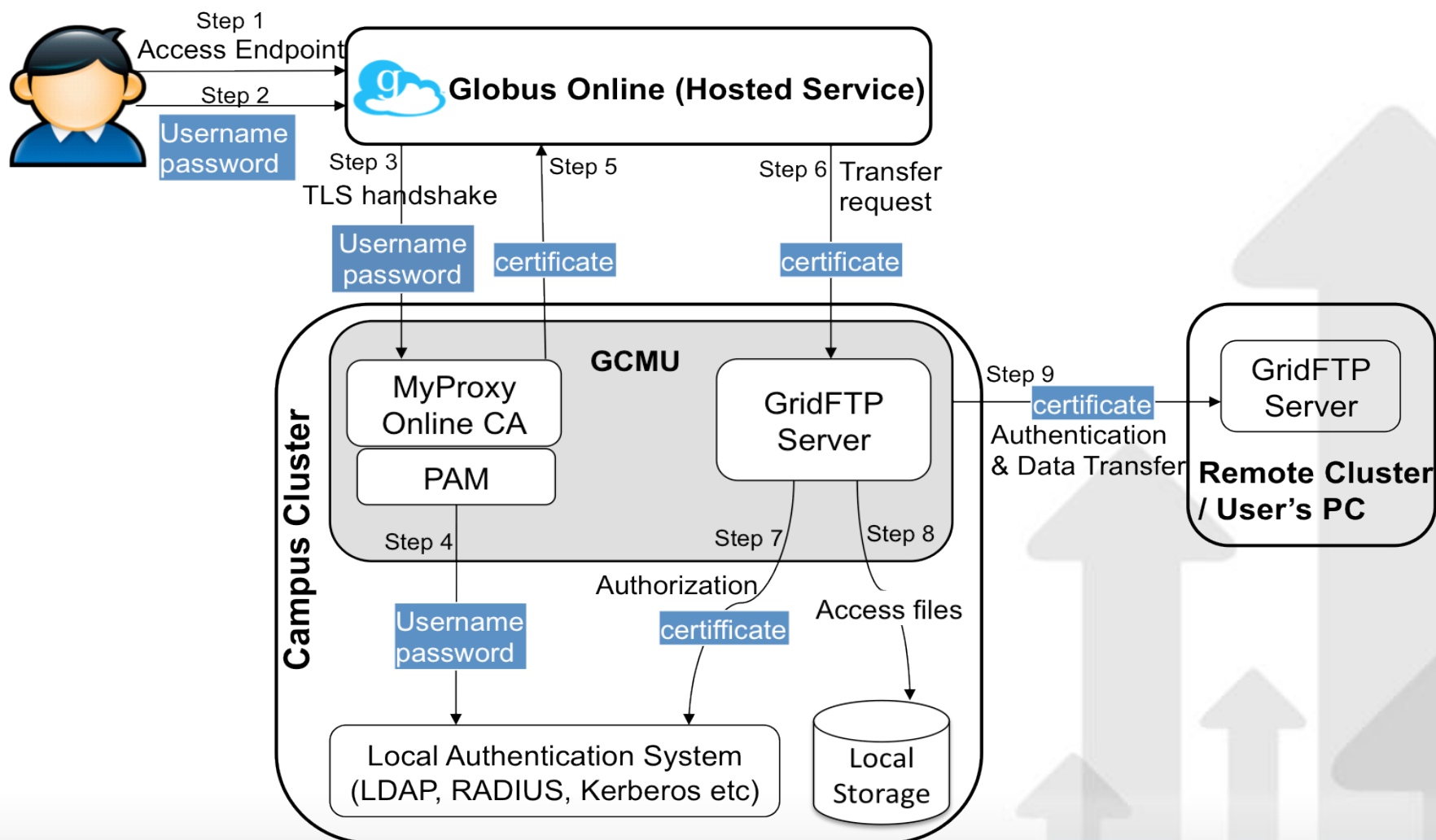


Data Channel Security Context (DCSC)





GO / GCMU Interaction





OAuth

- **Site passwords flow through Globus Online**
 - Globus Online does not store passwords
 - Just pass along to MyProxy servers at site
 - Still a security concern for some sites
- **OAuth**
 - Sites run a OAuth server
 - Users enter username and password only on a site's webpage
 - GO gets an X.509 credential via OAuth protocol



GCMU with OAuth

cd /opt

- **sudo wget <http://connect.globusonline.org/linux/stable/globusconnect-multiuser-1.2.0rc2.tgz>**
- **sudo tar xzf globusconnect-multiuser-1.2.0rc1.tgz**

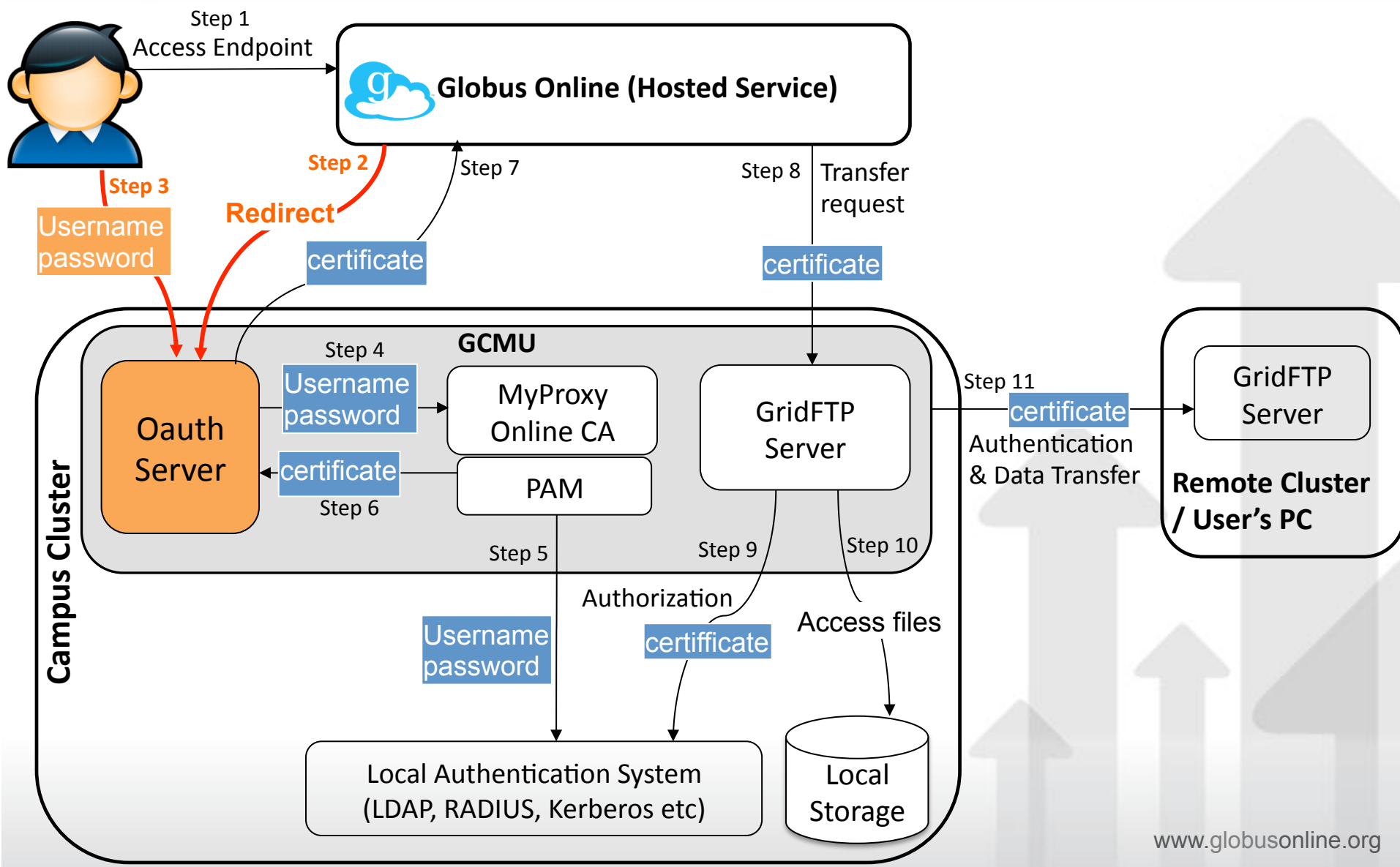
cd gcmu*

Sudo apt-get install python-flask

sudo ./install



Globus Connect Multi User





Details

- **GCMU may require firewall configuration**
 - Inbound ports: 2811, 7512, and 50000-51000
 - Outbound ports: 50000-51000
- **GCMU GridFTP cannot be (easily) used with other GridFTP clients besides GO**
 - GCMU uses a GO-issued cert, not a host cert
- **GridFTP has many configuration options**
 - E.g., Gridmap file can be used to allow other certs
 - Limit access to particular directories
- **MyProxy CA default proxy lifetime is 12 hours**



Single GridFTP server for different sets of users

- **GCMU allows you to serve different sets of users with a single GridFTP server**
- **For example, some XSEDE resource providers wants to setup a GridFTP server that caters to both XSEDE and non-XSEDE users**
 - **Setup a GCMU on top of XSEDE security infrastructure they already have**
 - **Separate GO endpoints e.g., xsede#ranger associated with xsede myproxy and tacc#ranger associated with local myproxy**
 - **Non-xsede users pick tacc#ranger and enter their tacc credentials to access it**